

Institutt for matematiske fag

Eksamensoppgave i **Løsningsskisse TMA4240 Statistikk**

Faglig kontakt under eksamen: Håkon Tjelmeland^a, Sara Martino^b

Tlf: ^a48 22 18 96, ^b99 40 33 30

Eksamensdato: 30. november 2017

Eksamenstid (fra–til): 09.00–13.00

Hjelpemiddelkode/Tillatte hjelpemidler: C

Tabeller og formler i statistikk, Akademika,
Bestemt, enkel kalkulator,
Gult stemplet A5-ark med egne håndskrevne notater.

Annen informasjon:

Alle svar skal begrunnes og besvarelsen skal inneholde naturlig mellomregning.

Målform/språk: bokmål

Antall sider: 11

Antall sider vedlegg: 0

Kontrollert av:

Informasjon om trykking av eksamensoppgave

Originalen er:

1-sidig ☐ 2-sidig ☒

sort/hvit ☒ farger ☐

skal ha flervalgskjema ☐

Dato

Sign

Oppgave 1

Det beskrevne stokastiske forsøket har $m = 6 \cdot 6 = 36$ mulige resultater, som kan organiseres i matrisa

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

der det første tallet angir antall øyne på første terning og den andre tallet angir antall øyne på den andre terningen. Av de mulige utfallene blir summen av antall øyne større enn eller lik 6 for de $g = 26$ utfallene angitt i rødt i tabellen over. Det er rimelig å anta en uniform sannsynlighetsmodell, og dermed får vi at

$$P(A) = \frac{g}{m} = \frac{26}{36} = \frac{13}{18} = \underline{\underline{0.7222}}.$$

Tolkningen av dette tallet er at dersom vi hadde gjentatt kast med to terninger uendelig mange ganger ville andelen av kastene som gav minst 6 øyne være $\frac{13}{18} = 0.7222$.

Hvis vi lar Y_1 være antall øyne på første terning og Y_2 være antall øyne på andre terning har vi at $X = Y_1 + Y_2$ slik at

$$E[X] = E[Y_1] + E[Y_2].$$

Dessuten har Y_1 og Y_2 samme sannsynlighetsfordeling slik at $E[Y_1] = E[Y_2]$. For en terning har vi dessuten at alle utfallene $y = 1, 2, \dots, 6$ har samme sannsynlighet $\frac{1}{6}$, slik at

$$E[Y_1] = E[Y_2] = \sum_{y=1}^6 y \cdot \frac{1}{6} = \frac{1}{6} \cdot (1 + 2 + 3 + 4 + 5 + 6) = \frac{21}{6}.$$

Vi får dermed at

$$E[X] = \frac{21}{6} + \frac{21}{6} = \frac{21}{3} = \underline{\underline{7}}.$$

Tolkningen av dette tallet er at dersom vi hadde gjentatt kast med to terninger uendelig mange ganger ville vi i gjennomsnitt fått 7 øyne på hvert kast.

Oppgave 2 Oppfølgingsavdelingen

- a) La V_A betegne antall besøk en pasient fra avdeling A har til oppfølgingsavdelingen. Det er da angitt at

$$V_A \sim p(v_A; \mu_A = 1.4) = \frac{1.4^{v_A}}{v_A!} e^{-1.4} \text{ for } v_A = 0, 1, 2, \dots$$

Vi får dermed at

$$\underline{\underline{P(V_A = 0) = \frac{1.4^0}{0!} e^{-1.4} = e^{-1.4} = 0.2466,}}$$

og

$$\begin{aligned} P(V_A > 2 | V_A \geq 1) &= \frac{P(V_A > 2 \cap V_A \geq 1)}{P(V_A \geq 1)} = \frac{P(V_A > 2)}{P(V_A \geq 1)} \\ &= \frac{1 - P(V_A \leq 2)}{1 - P(V_A = 0)} \\ &= \frac{1 - P(V_A = 0) - P(V_A = 1) - P(V_A = 2)}{1 - P(V_A = 0)} \\ &= \frac{1 - 0.2466 - \frac{1.4^1}{1!} e^{-1.4} - \frac{1.4^2}{2!} e^{-1.4}}{1 - 0.2466} \\ &= \underline{\underline{0.221.}} \end{aligned}$$

La A betegne hendelsen at en tilfeldig valgt pasient på oppfølgingsavdelingen kommer fra avdeling A og la B betegne at pasienten kommer fra avdeling B. Vi har da opplyst at $P(A) = 0.66$ og $P(B) = 1 - P(A) = 0.34$. La V betegne antall besøk til oppfølgingsavdelingen for en tilfeldig valgt pasient. Ved å bruke setningen om total sannsynlighet får vi dermed at

$$\begin{aligned} P(V = 0) &= P(V = 0 | A)P(A) + P(V = 0 | B)P(B) \\ &= P(V_A = 0)P(A) + P(V_B = 0)P(B) \\ &= 0.2466 \cdot 0.66 + \frac{0.81^0}{0!} e^{-0.81} \cdot 0.34 = \underline{\underline{0.3140.}} \end{aligned}$$

- b) La X_A og X_B være antall pasienter operert ved henholdsvis avdeling A og B som har ingen besøk hos oppfølgingsavdelingen. Siden hver pasient operert ved avdeling A har sannsynlighet $p_A = P(V_A = 0) = 0.2466$ for å ha ingen besøk til oppfølgingsavdelingen, uavhengig av hverandre, blir X_A binomisk fordelt,

$$X_A \sim b(x_A; n_A = 16302, p_A = 0.2466).$$

Tilsvarende blir sannsynligheten for at en pasient operert ved avdeling B har ingen besøk ved oppfølgingsavdelingen $p_B = \frac{0.81^0}{0!} e^{-0.81} = 0.4449$ og

$$X_B \sim b(x_B; n_B = 8398, p_B = 0.4449).$$

Ved å benytte at $X = X_A + X_B$ og kjent formel for forventingsverdien i en binomisk fordeling får vi dermed at

$$\begin{aligned} E[X] &= E[X_A + X_B] = E[X_A] + E[X_B] = n_A p_A + n_B p_B \\ &= 16302 \cdot 0.2466 + 8398 \cdot 0.4449 = \underline{\underline{7756.343}}. \end{aligned}$$

Ved å benytte at X_A og X_B er uavhengige og kjent formel for variansen i en binomisk fordeling får vi

$$\begin{aligned} \text{Var}[X] &= \text{Var}[X_A + X_B] = \text{Var}[X_A] + \text{Var}[X_B] \\ &= n_A p_A (1 - p_A) + n_B p_B (1 - p_B) \\ &= 16302 \cdot 0.2466 \cdot (1 - 0.2466) + 8398 \cdot 0.4449 \cdot (1 - 0.4449) \\ &= 5102.727, \end{aligned}$$

slik at

$$\text{SD}[X] = \sqrt{\text{Var}[X]} = \sqrt{5102.727} = \underline{\underline{71.4334}}.$$

Vi vet at en sum av uavhengige poissonfordelte variabler er poissonfordelt. Siden antall besøk til oppfølgingsavdelingen for hver pasient er poissonfordelt og antall besøk for ulike pasienter er uavhengige blir dermed totalt antall besøk til oppfølgingsavdelingen poissonfordelt.

La Y_{Ai} for $i = 1, 2, \dots, n_A$ være antall besøk for pasient nummer i fra avdeling A og la Y_{Bi} for $i = 1, 2, \dots, n_B$ være antall besøk for pasient nummer i fra avdeling B. Da har vi $E[Y_{Ai}] = \mu_A = 1.4$ og $E[Y_{Bi}] = \mu_B = 0.81$, og dermed

$$\begin{aligned} E[Y] &= E\left[\sum_{i=1}^{n_A} Y_{Ai} + \sum_{i=1}^{n_B} Y_{Bi}\right] = \sum_{i=1}^{n_A} E[Y_{Ai}] + \sum_{i=1}^{n_B} E[Y_{Bi}] \\ &= \sum_{i=1}^{n_A} 1.4 + \sum_{i=1}^{n_B} 0.81 = 1.4 n_A + 0.81 n_B = 1.4 \cdot 16302 + 0.81 \cdot 8398 = \underline{\underline{29625.18}}. \end{aligned}$$

Siden vi vet at Y er poissonfordelt vet vi at $\text{Var}[Y] = E[Y]$ slik at vi får

$$\text{SD}[Y] = \sqrt{\text{Var}[Y]} = \sqrt{29625.18} = \underline{\underline{172.1197}}.$$

Oppgave 3 Levetid til en ny type mekaniske komponentera) Innsatt $z = 1$ får vi

$$f(y) = \frac{2y}{\theta^2} e^{-\frac{y^2}{\theta^2}} \quad \text{for } y > 0.$$

For $y > 0$ får vi

$$\begin{aligned} F(y) &= \int_{-\infty}^y f(y) dy = \int_0^y \frac{2y}{\theta^2} e^{-\frac{y^2}{\theta^2}} dy \\ &= \left[-e^{-\frac{y^2}{\theta^2}} \right]_0^y = -e^{-\frac{y^2}{\theta^2}} - (-e^0) = 1 - e^{-\frac{y^2}{\theta^2}}. \end{aligned}$$

Dermed har vi at

$$F(y) = \begin{cases} 1 - e^{-\frac{y^2}{\theta^2}} & \text{for } y > 0, \\ 0 & \text{ellers.} \end{cases}$$

Fra dette finner vi så at

$$P(Y > 500) = 1 - P(\leq 500) = 1 - F(500) = 1 - \left(1 - e^{-\frac{500^2}{1000^2}} \right) = e^{-\frac{500^2}{1000^2}} = \underline{\underline{0.7788}}.$$

Medianen m er gitt fra ligningen $F(m) = 0.5$, som gir at

$$\begin{aligned} 1 - e^{-\frac{m^2}{\theta^2}} &= 0.5 \\ e^{-\frac{m^2}{\theta^2}} &= 0.5 \\ -\frac{m^2}{\theta^2} &= \ln(0.5) \\ m^2 &= -\theta^2 \ln(0.5) \\ m &= \theta \sqrt{-\ln(0.5)} \\ m &= 1000 \sqrt{-\ln(0.5)} = \underline{\underline{832.5546}}. \end{aligned}$$

b) Rimelighetsfunksjonen blir

$$L(\theta) = f(y_1, y_2, \dots, y_n; \theta) = \prod_{i=1}^n f(y_i; \theta) = \prod_{i=1}^n \left[\frac{2y_i z_i^2}{\theta^2} e^{-\frac{y_i^2 z_i^2}{\theta^2}} \right]$$

Log-rimelighetsfunksjonen blir dermed

$$\begin{aligned} l(\theta) &= \ln L(\theta) = \sum_{i=1}^n \left[\ln 2 + \ln(y_i) + 2 \ln(z_i) - 2 \ln \theta - \frac{y_i^2 z_i^2}{\theta^2} \right] \\ &= n \ln 2 + \sum_{i=1}^n \ln(y_i) + 2 \sum_{i=1}^n \ln(z_i) - 2n \ln \theta - \frac{1}{\theta^2} \sum_{i=1}^n y_i^2 z_i^2. \end{aligned}$$

Deriverer $l(\theta)$ med hensyn på θ ,

$$l'(\theta) = -\frac{2n}{\theta} + \frac{2}{\theta^3} \sum_{i=1}^n y_i^2 z_i^2 = -\frac{2}{\theta} \left[n - \frac{1}{\theta^2} \sum_{i=1}^n y_i^2 z_i^2 \right].$$

Finner verdien som maksimerer $l(\theta)$ ved å sette $l'(\theta) = 0$, som gir

$$\begin{aligned} n &= \frac{1}{\theta^2} \sum_{i=1}^n y_i^2 z_i^2 \\ \theta^2 &= \frac{1}{n} \sum_{i=1}^n y_i^2 z_i^2 \\ \theta &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i z_i)^2}. \end{aligned}$$

SME blir dermed

$$\hat{\theta} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i z_i)^2}.$$

Innsatt tall fra tabell 1 får man at estimatet blir

$$\hat{\theta} = \sqrt{\frac{1}{10} \cdot 6736616} = \underline{\underline{820.77}}.$$

c) Den en-entydige transformasjonen mellom y_i og u_i er

$$u_i = \frac{2y_i^2 z_i^2}{\theta^2} \Leftrightarrow y_i = \sqrt{\frac{\theta^2 u_i}{2z_i^2}}.$$

Siden vi kun er interessert i $y_i > 0$ og $u_i > 0$ er transformasjonen en-entydig. Den deriverte av y_i ved hensyn på u_i blir

$$\frac{dy_i}{du_i} = \frac{1}{2\sqrt{\frac{\theta^2 u_i}{2z_i^2}}} \cdot \frac{\theta^2}{2z_i^2}$$

Da $u_i > 0$ blir den deriverte alltid positiv. For $y_i, u_i > 0$ gir transformasjonsformelen da

$$\begin{aligned} f_{U_i}(u_i) &= f_{Y_i}(y_i) \cdot \left| \frac{dy_i}{du_i} \right| \\ &= \frac{2\sqrt{\frac{\theta^2 u_i}{2z_i^2}} z_i^2}{\theta^2} e^{-\frac{\left(\sqrt{\frac{\theta^2 u_i}{2z_i^2}}\right)^2 z_i^2}{\theta^2}} \cdot \frac{1}{2\sqrt{\frac{\theta^2 u_i}{2z_i^2}}} \cdot \frac{\theta^2}{2z_i^2} \\ &= \frac{1}{2} e^{-\frac{u_i}{2}}. \end{aligned}$$

Sannsynlighetstettheten til en χ^2_ν -fordeling er for $x > 0$

$$f(x) = \frac{1}{2^{\frac{\nu}{2}} \Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} \quad \text{for } x > 0.$$

Innsatt $\nu = 2$ blir denne

$$f(x) = \frac{1}{2^1 \Gamma(1)} x^{1-1} e^{-\frac{x}{2}} = \frac{1}{2} e^{-\frac{x}{2}},$$

som vi ser er identisk med sannsynlighetstettheten til U_i . Vi har dermed at $\underline{U_i \sim \chi^2_2}$.

Ved å sette inn uttrykket vi har for $\hat{\theta}$ får vi

$$\frac{2n\hat{\theta}^2}{\theta^2} = \frac{2n \cdot \frac{1}{n} \sum_{i=1}^n (Y_i z_i)^2}{\theta^2} = \sum_{i=1}^n \frac{2Y_i^2 z_i^2}{\theta^2} = \sum_{i=1}^n U_i.$$

Siden Y_i 'ene er antatt uavhengige blir også U_i 'ene uavhengig, og vi vet at en sum av uavhengige $\chi^2_{\nu_i}$ -fordelte variabler er χ^2_ν -fordelt med $\nu = \sum_{i=1}^n \nu_i$. Siden $U_i \sim \chi^2_2$ får vi dermed at $\sum_{i=1}^n U_i$ er χ^2 -fordelt med $\sum_{i=1}^n 2 = 2n$ frihetsgrader, dvs

$$\underline{\underline{\frac{2n\hat{\theta}^2}{\theta^2} \sim \chi^2_{2n}}}.$$

d) Siden

$$\frac{2n\hat{\theta}^2}{\theta^2} \sim \chi^2_{2n}$$

har vi at

$$P\left(x_{1-\frac{\alpha}{2}, 2n}^2 \leq \frac{2n\hat{\theta}^2}{\theta^2} \leq x_{\frac{\alpha}{2}, 2n}^2\right) = 1 - \alpha.$$

For å omskrive dette uttrykket løser vi hver av de to ulikhetene med hensyn på θ , starter med den første ulikheten,

$$x_{1-\frac{\alpha}{2}, 2n}^2 \leq \frac{2n\hat{\theta}^2}{\theta^2} \Leftrightarrow \theta^2 \leq \frac{2n\hat{\theta}^2}{x_{1-\frac{\alpha}{2}, 2n}^2} \Leftrightarrow \theta \leq \hat{\theta} \sqrt{\frac{2n}{x_{1-\frac{\alpha}{2}, 2n}^2}},$$

siden $\theta > 0$ og $\hat{\theta} > 0$. For den andre ulikheten får vi tilsvarende

$$\frac{2n\hat{\theta}^2}{\theta^2} \leq x_{\frac{\alpha}{2}, 2n}^2 \Leftrightarrow \frac{2n\hat{\theta}^2}{x_{\frac{\alpha}{2}, 2n}^2} \leq \theta^2 \Leftrightarrow \hat{\theta} \sqrt{\frac{2n}{x_{\frac{\alpha}{2}, 2n}^2}} \leq \theta.$$

Ved å sette de to ulikhetene sammen igjen med θ i midten får vi dermed

$$P\left(\hat{\theta}\sqrt{\frac{2n}{x_{\frac{\alpha}{2}, 2n}^2}} \leq \theta \leq \hat{\theta}\sqrt{\frac{2n}{x_{1-\frac{\alpha}{2}, 2n}^2}}\right) = 1 - \alpha.$$

Et $(1 - \alpha) \cdot 100\%$ -konfidensintervall for θ blir dermed

$$\left[\hat{\theta}\sqrt{\frac{2n}{x_{\frac{\alpha}{2}, 2n}^2}}, \hat{\theta}\sqrt{\frac{2n}{x_{1-\frac{\alpha}{2}, 2n}^2}} \right].$$

Med $\alpha = 0.05$ får vi fra tabell over kvantiler i χ^2 -fordelinger at $x_{1-\frac{\alpha}{2}, 2n}^2 = x_{0.0975, 20}^2 = 9.591$ og $x_{\frac{\alpha}{2}, 2n}^2 = x_{0.025, 20}^2 = 34.170$. Innsatt tall fra tabell 1 får vi at

$$\hat{\theta} = \sqrt{\frac{1}{n} \sum_{i=1}^n y_i z_i} = \sqrt{\frac{1}{10} \cdot 6736616} = 820.7689,$$

slik at konfidensintervallet blir

$$\left[820.7689 \cdot \sqrt{\frac{20}{34.170}}, 820.7689 \cdot \sqrt{\frac{20}{9.591}} \right] = \underline{\underline{[627.9331, 1185.234]}}.$$

Oppgave 4 Infeksjon etter operasjon

- a) Sentralgrenseteoremet sier at dersom $X_1, X_2, \dots, X_n$ er uavhengige og identiske fordelte med en forventingsverdi $E[X_i] = \mu$ og en varians $\text{Var}[X_i] = \sigma^2$ vil sannsynlighetsfordelingen til

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$$

konvergere mot en standard-normalfordeling når $n \rightarrow \infty$.

Når n er stor betyr dette at

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \approx n(z; 0, 1).$$

Dessuten har vi at

$$\bar{X} = \mu + Z\sqrt{\frac{\sigma^2}{n}},$$

dvs $\bar{X}$ er en lineær funksjon av Z . Dermed er også $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ tilnærmet normalfordelt.

Hvis vi så ser på situasjonen beskrevet i oppgaven, og fokuserer først på pasientene som er operert før omorganiseringen, kan vi definere

$$X_{Fi} = \begin{cases} 1 & \text{hvis pasient nr } i \text{ i perioden før omorganiseringen får infeksjon,} \\ 0 & \text{ellers.} \end{cases}$$

Vi har da at $\mu = E[X_{Fi}] = p_F$ og $\sigma^2 = \text{Var}[X_{Fi}] = E[X_{Fi}^2] - E[X_{Fi}]^2 = p_F - p_F^2$. Fra det generelle resultatet over har vi da at $\hat{p}_F = \frac{X_F}{n_F} = \frac{1}{n} \sum_{i=1}^{n_F} X_{Fi}$ er tilnærmet normalfordelt.

At også $\hat{p}_E$ er tilnærmet normalfordelt begrunnes helt tilsvarende som for $\hat{p}_F$.

Siden forskjellige pasienter får infeksjon uavhengig av hverandre blir X_E og X_F uavhengige, som i sin tur gir at $\hat{p}_F$ og $\hat{p}_E$ blir uavhengige. Dermed er $\hat{p}_E - \hat{p}_F$ (tilnærmet) normalfordelt fordi den er en lineær funksjon av uavhengige og (tilnærmet) normalfordelte variabler.

Ved å benytte regneregler for forventingsverdi og at $X_F \sim b(x_F; n_F, p_F)$ og $X_E \sim b(x_E; n_E, p_E)$ får vi at

$$\begin{aligned} E[\hat{p}_E - \hat{p}_F] &= E[\hat{p}_E] - E[\hat{p}_F] = E\left[\frac{X_E}{n_E}\right] - E\left[\frac{X_F}{n_F}\right] \\ &= \frac{1}{n_E} E[X_E] - \frac{1}{n_F} E[X_F] = \frac{1}{n_E} \cdot n_E p_E - \frac{1}{n_F} \cdot n_F p_F \\ &= \underline{\underline{p_E - p_F}}. \end{aligned}$$

Ved å bruke regneregler for varians og at X_F og X_E er uavhengige får vi

$$\begin{aligned} \text{Var}[\hat{p}_E - \hat{p}_F] &= \text{Var}\left[\frac{X_E}{n_E} - \frac{X_F}{n_F}\right] = \text{Var}\left[\frac{X_E}{n_E}\right] + (-1)^2 \cdot \text{Var}\left[\frac{X_F}{n_F}\right] \\ &= \frac{1}{n_E^2} \text{Var}[X_E] + \frac{1}{n_F^2} \text{Var}[X_F] \\ &= \frac{1}{n_E^2} \cdot n_E p_E (1 - p_E) + \frac{1}{n_F^2} \cdot n_F p_F (1 - p_F) \\ &= \underline{\underline{\frac{p_E(1 - p_E)}{n_E} + \frac{p_F(1 - p_F)}{n_F}}}. \end{aligned}$$

- b) Vi ønsker å kunne konkludere med at omorganiseringen har vært vellykket, dvs vi ønsker å kunne konkludere med at andelen som får infeksjon er mindre

etter omorganiseringen. Vi må da som H_1 velge at $p_E < p_F \Leftrightarrow p_E - p_F < 0$.
Skal følgende teste

$$\underline{\underline{H_0 : p_E - p_F = 0 \quad \text{mot} \quad H_1 : p_E - p_F < 0.}}$$

Ved å ta utgangspunkt i resultatene i a) og standardisere $\hat{p}_E - \hat{p}_F$ får vi at

$$\frac{(\hat{p}_E - \hat{p}_F) - (p_E - p_F)}{\sqrt{\frac{p_E(1-p_E)}{n_E} + \frac{p_F(1-p_F)}{n_F}}} \approx n(z; 0, 1).$$

Når H_0 er riktig får vi dermed også at

$$\frac{\hat{p}_E - \hat{p}_F}{\sqrt{\frac{p_E(1-p_E)}{n_E} + \frac{p_F(1-p_F)}{n_F}}} \approx n(z; 0, 1). \quad (1)$$

Denne størrelsen kan vi ikke benytte som testobservator siden p_E og p_F som inngår nevneren er ukjente. Når H_0 er riktig er $p_E = p_F$ og vi lar p være denne felles verdien. Når H_0 er riktig vil $X_E + X_F \sim b(x; n_E + n_F, p)$ slik at vi kan estimere p ved

$$\hat{p} = \frac{X_E + X_F}{n_E + n_F}.$$

Siden $n_E + n_F$ er stor blir $\hat{p} \approx p = p_E = p_F$, slik at ved å erstatte p_E og p_F med $\hat{p}$ i (1) får vi en størrelse vi kan benytte som testobservator,

$$\underline{\underline{Z = \frac{\hat{p}_E - \hat{p}_F}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_E} + \frac{1}{n_F}\right)}} \approx n(z; 0, 1) \quad \text{når } H_0 \text{ er riktig.}}}$$

For å finne p -verdien må vi først finne observert verdi for testobservatoren.

Innsatt observerte tall får vi $\hat{p}_E = \frac{x_E}{n_E} = \frac{135}{1919}$, $\hat{p}_F = \frac{x_F}{n_F} = \frac{186}{2021}$,

$$\hat{p} = \frac{x_E + x_F}{n_E + n_F} = \frac{186 + 135}{2021 + 1919} = \frac{321}{3940},$$

og observert verdi for testobservatoren blir dermed

$$\begin{aligned} z_{\text{obs}} &= \frac{\hat{p}_E - \hat{p}_F}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_E} + \frac{1}{n_F}\right)}} \\ &= \frac{\frac{135}{1919} - \frac{186}{2021}}{\sqrt{\frac{321}{3940}\left(1 - \frac{321}{3940}\right)\left(\frac{1}{1919} + \frac{1}{2021}\right)}} = -2.49 \end{aligned}$$

Man vil naturlig forkaste H_0 dersom Z er liten nok, så p -verdien blir

$$p = P(Z \leq z_{\text{obs}} | H_0) = P(Z \leq -2.49 | H_0) = \underline{\underline{0.0064}}.$$

Sannsynligheten for å observere det vi har observert eller noe mer ekstremt dersom H_0 er riktig er altså så lav som 0.0064. Det er følgelig rimelig å forkaste H_0 , dvs konkludere med at omorganiseringen har vært vellykket.

- c) La Y være antall av pasientene som blir operert neste år som får infeksjon. Vi har da at Y er uavhengig av X_E og at $Y \sim b(y; m, p_E)$. Ved å argumentere tilsvarende som i a) får vi at

$$Y - \frac{m}{n_E} X_E$$

er tilnærmet normalfordelt med

$$E \left[Y - \frac{m}{n_E} X_E \right] = mp_E - \frac{m}{n_E} \cdot n_E p_E = 0$$

og

$$\begin{aligned} \text{Var} \left[Y - \frac{m}{n_E} X_E \right] &= mp_E(1 - p_E) + \left(\frac{m}{n_E} \right)^2 n_E p_E(1 - p_E) \\ &= m \left(1 + \frac{m}{n_E} \right) p_E(1 - p_E). \end{aligned}$$

Ved å standardisere får vi dermed

$$\frac{Y - \frac{m}{n_E} X_E}{\sqrt{m \left(1 + \frac{m}{n_E} \right) p_E(1 - p_E)}} \approx n(z; 0, 1).$$

Erstatter så p_E i dette uttrykket med $\hat{p}_E = \frac{X_E}{n_E}$ og bruker at $\hat{p}_E \approx p_E$ siden n_E er stor, til å konkludere at

$$\frac{Y - \frac{m}{n_E} X_E}{\sqrt{m \left(1 + \frac{m}{n_E} \right) \frac{X_E}{n_E} \left(1 - \frac{X_E}{n_E} \right)}} \approx n(z; 0, 1).$$

Dermed har vi at

$$P \left(-z_{\frac{\alpha}{2}} \leq \frac{Y - \frac{m}{n_E} X_E}{\sqrt{m \left(1 + \frac{m}{n_E} \right) \frac{X_E}{n_E} \left(1 - \frac{X_E}{n_E} \right)}} \leq z_{\frac{\alpha}{2}} \right) \approx 1 - \alpha.$$

Løser hver av ulikhetene i dette uttrykket med hensyn på Y . Den første ulikheten gir

$$Y \geq \frac{m}{n_E} X_E - z_{\frac{\alpha}{2}} \sqrt{m \left(1 + \frac{m}{n_E} \right) \frac{X_E}{n_E} \left(1 - \frac{X_E}{n_E} \right)},$$

og den andre gir

$$Y \leq \frac{m}{n_E} X_E + z_{\frac{\alpha}{2}} \sqrt{m \left(1 + \frac{m}{n_E}\right) \frac{X_E}{n_E} \left(1 - \frac{X_E}{n_E}\right)}.$$

Ved å sette de to ulikhetene sammen igjen med Y i midten har man dermed at

$$\begin{aligned} P \left(\frac{m}{n_E} X_E - z_{\frac{\alpha}{2}} \sqrt{m \left(1 + \frac{m}{n_E}\right) \frac{X_E}{n_E} \left(1 - \frac{X_E}{n_E}\right)} \leq Y \right. \\ \left. \leq \frac{m}{n_E} X_E + z_{\frac{\alpha}{2}} \sqrt{m \left(1 + \frac{m}{n_E}\right) \frac{X_E}{n_E} \left(1 - \frac{X_E}{n_E}\right)} \right) \approx 1 - \alpha. \end{aligned}$$

Et $(1 - \alpha) \cdot 100\%$ -prediksjonsintervall for Y er dermed

$$\left[\frac{m}{n_E} X_E - z_{\frac{\alpha}{2}} \sqrt{m \left(1 + \frac{m}{n_E}\right) \frac{X_E}{n_E} \left(1 - \frac{X_E}{n_E}\right)}, \frac{m}{n_E} X_E + z_{\frac{\alpha}{2}} \sqrt{m \left(1 + \frac{m}{n_E}\right) \frac{X_E}{n_E} \left(1 - \frac{X_E}{n_E}\right)} \right]$$

Innsatt $\alpha = 0.1$, og dermed $z_{\frac{\alpha}{2}} = z_{0.05} = 1.645$, og observerte verdier blir prediksjonsintervallet

$$\begin{aligned} & \left[\frac{2000}{1919} \cdot 135 - 1.645 \sqrt{2000 \left(1 + \frac{2000}{1919}\right) \frac{135}{1919} \left(1 - \frac{135}{1919}\right)}, \right. \\ & \left. \frac{2000}{1919} \cdot 135 + 1.645 \sqrt{2000 \left(1 + \frac{2000}{1919}\right) \frac{135}{1919} \left(1 - \frac{135}{1919}\right)} \right] \\ & \underline{\underline{[113.81, 167.58]}}. \end{aligned}$$